DOCUMENT RESUME

ED 210 288 TM 810 927

AUTHOR Apling, Richard; Bryk, Anthony
TITLE Evaluation Approaches: A Focus on Improving Early Childhood Title I Programs.
INSTITUTION Huron Inst., Cambridge, Mass.
SPONS AGENCY Department of Education, Washington, D.C.
PUB DATE Dec 80
NOTE 66p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Early Childhood Education; *Evaluation Methods; Informal Assessment; *Program Evaluation; Program Improvement; Self Evaluation (Groups)
IDENTIFIERS *Elementary Secondary Education Act Title I; External Evaluation; Qualitative Analysis; Quantitative Analysis

ABSTRACT

Prepared as part of a United Stated Education Department project on evaluation in early childhood Title I (ECT-I) programs, this is one of a series of resource books developed in response to concerns expressed by state and local personnel about early childhood Title I programs. By considering: (1) who will use the evaluation results; (2) what kinds of information are users likely to find most helpful; (3) in what ways might this information aid in program improvement; and (4) are the potential benefits substantial enough to justify the cost and effort of evaluation. The purpose is not to be a comprehensive technical manual, but rather to help local school personnel identify issues that might merit further examination and to guide the choice of suitable evaluation strategies to address those issues. From the initial program review methods, citing the human resource requirement and procedures for conducting self-study and outside review, the document outlines quantative methods (such as structured process evaluation, product evaluation and process-product evaluation) and qualitative methods (investigative evaluation, ethnographic evaluation, and documentation) and lists reasons to choose a method and outlines how to implement the plan. The underlying theme would have the reader use this as a springboard for beginning evaluation. (CE)

EVALUATION APPROACHES:

A FOCUS ON IMPROVING

EARLY CHILDHOOD TITLE I PROGRAMS

December 1980

Richard Apling and Anthony Bryk
The Huron Institute
123 Mt. Auburn Street
Cambridge, Massachusetts 02138

FOREWORD

This booklet has been prepared as part of a project sponsored by the United States Education Department (USED) on evaluation in early childhood Title I (ECT-I) programs. It is one of a series of resource books developed in response to concerns expressed by state and local personnel about early childhood Title I programs. The series describes an array of diverse evaluation activities and outlines how each of these might contribute to improving local programs. The series revolves around a set of questions:

- Who will use the evaluation results?
- What kinds of information are users likely to find most helpful?
- In what ways might this information aid in program improvement?
- Are the potential benefits substantial enough to justify the cost and effort of evaluation?

Together, the resource bocks address a range of issues relevant to the evaluation of early childhood programs for educationally disadvantaged children. The series comprises the following volumes:

- Evaluating Title I Early Childhood Programs: An Overview
- Assessment in Early Childhood Education
- Short-Term Impact Evaluation of Early Childhood Title I Programs
- An Introduction to the Value-Added Model and Its Use in Short-Term Impact Assessment
- Evaluation Approaches: A Focus on Improving Early Childhood Title I Programs
- Longitudinal Evaluation Systems for Early Childhood Title I Programs
- Evaluating Title I Parent Education Programs

The development of this series follows extensive field work on ECT-I programs (Yurchak & Bryk, 1979). In the course of that research, we

identified a number of concerns that SEA and LEA officials had about ECT-I programs, and the kinds of information that might be helpful in addressing them. Each resource book in the series thus deals with a specific concern or set of concerns. The books and the evaluation approaches they describe do not, however, constitute a comprehensive evaluation system to be uniformly applied by all. Our feasibility analysis (Bryk, Apling, & Mathews, 1978) indicated that such a system could not efficiently respond to the specific issues of interest in any single district at any given time. Rather, LEA personnel might wish to draw upon one or more of the approaches we describe, tailoring their effort to fit the particular problem confronting them.

Finally, the resource books are not comprehensive technical manuals. Their purpose is to help local school personnel identify issues that might merit further examination and to guide the choice of suitable evaluation strategies to address those issues. Additional information and assistance in using the various evaluation strategies are available in the more technical publications cited at the end of each volume, and from the Technical Assistance Centers in the ten national regions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. BACKGROUND AND RATIONALE

Program evaluation has been a requirement under Title I since its inception. Over the years, most evaluations have concentrated on the accountability question: Has the program achieved its goals? This information, required for local educational agencies' (LEAs) annual funding applications, is potentially useful to school administrators, school boards, and Parent Advisory Councils (PACs) in examining the overall effectiveness of school programs. The Department of Education (ED) proposed evaluation models for grades 2 through 12, and comparable procedures for early childhood programs have been designed to generate such information (see Tallmadge & Wood, 1978, and the resource book Short-Term Impact Evaluations of Early Childhood Title I Programs by Haney, 1980).

Recently, attention has begun to focus on how the results of evaluation might lead to improved program practice. This is reflected in the Education Amendments of 1978, which stipulate that Title I evaluations should provide LEA staff with information for planning and improving their projects and activities. In the Huron Institute's visits to early childhood Title I (ECT-I) programs across the country, we found a number of school districts that had developed their own evaluation activities to address local concerns about improving ECT-I programs, and other districts that were interested but did not know quite how to proceed.

Unfortunately, the evaluation models for grades 2 through 12 and the early childhood equivalents were designed primarily to generate data about overall program effectiveness. As a result, they often will not provide sufficient detail to LEA staff concerned about where a program may be faltering, why problems are occurring, and how improvements can be made.

This resource book is a first response to that need. We survey here a wide range of evaluation techniques that can generate quite extensive information about programs, which can be helpful in suggesting ways to improve. In choosing techniques to discuss, we have drawn both on methods that were used in ECT-I programs we visited and on other promising approaches. Some of these methods can also produce information useful for school district policy formulation and accountability, or data about individual children that teachers can use to make instructional decisions; we present them here because they seem to us especially useful to local staff as they attempt to improve their programs. While all the methods are intended to inform efforts to improve programs, each addresses somewhat different questions; thus their usefulness will vary from context to context.

In describing these evaluation methods, we try to suggest the kinds of question that each can address, the kinds of information that it might produce, and the resources and personnel commitments that it requires. Each of the methods we discuss, however, is not so much a precise and well defined set of activities as a general approach to inquiry. All have considerable merit and no one approach is unequivocally best. For an evaluation to generate useful information for program planning and improvement, it must be tailored to local concerns and the nature of the particular program. Although we introduce each as a separate method, actual evaluations might consist of a variant on one approach or some combination of approaches. These may be used to look at an ECT-I program in its entirety, or perhaps at just one program component or particular feature. In short, the possibilities are enormous. Our modest goal for this resource book, then, is to suggest some of the possibilities, but not to fully map the terrain.

LEA program staff will usually find it necessary to draw on evaluation expertise in designing and implementing such a study. While some LEAs have an evaluation staff to support these efforts, many may want to seek outside help, for example from the Title I Technical Assistance Centers (TACs) or from local evaluation consultants. This document is not a substitute for professional assistance. Rather, we intend it as an idea book for local program staff on how evaluation can help them to plan and improve programs.

We have organized the evaluation methods described in this resource book into three basic types:

- Program review methods such as self-study and outside review
- Quantitative social science research methods that use student test data and program descriptive information
- Methods that are generally referred to as naturalistic or qualitative inquiry.

The following three chapters describe these three types of methods and illustrate them with short examples. For each type, we discuss human resource requirements, the procedures to be used, and some situations where the methods might be most useful. In the concluding chapter, we summarize the alternatives presented and raise a number of issues that LEAs should consider about the process of developing and implementing an evaluation plan.

II. PROGRAM REVIEW METHODS

Both self-study and outside review are ways of examining ECT-I programs that feature the full involvement in the evaluation process of teachers and other program staff from the school system. ECT-I staff can examine both general program features such as program goals, parent participation, and staff qualifications, and specific activities such as teaching practices and selection procedures. Although the two methods are similar, the audiences for the evaluations may differ depending on whether the program staff or an outside team evaluates the program.

In a self-study, staff members assess their own program. Working in committees, they identify strengths and weaknesses and make recommendations for improvement. Not only do ECT-I staff members conduct the evaluation, but they often become the chief audience for the results. Although the school board, local administrators, and parent groups might receive the final report on the self-study, it is the staff that is most likely to use the results of the evaluation to improve program practice.

In an outside review, a visiting committee, usually consisting of staff from nearby ECT-I programs and early childhood education specialists, examines a program's goals, resources, materials, procedures, personnel qualifications, and facilities. As in self-study, the committee looks for strengths and weaknesses in each area and recommends improvements. Unlike self-study, which aims primarily at self-examination and self-improvement, outside review may also treat accountability considerations, for example in program accreditation. Thus, while ECT-I staff members are often the main users of the results of self-study, outside review may be aimed more at parents, program administrators, and school boards.

AN EXAMPLE: OUTSIDE REVIEW OF A SPECIAL EDUCATION PROGRAM

A state director of special education and his staff initiated an outside review to identify exemplary projects and to provide local program staffs with suggestions for improving their practice. Each visiting team was composed of representatives of the state division of special education, special education administrators and teachers, regular education staff, and parents. In looking at the special education program in one community, the team examined the following areas:

- The preschool component
- Kindergarten screening
- Staff development
- Program delivery
- Public communications
- Transportation
- Physical facilities.

The team reviewed data in student records, interviewed staff, and had parents fill out questionnaires. For each area, the team identified commendable practices, pointed out problems, and recommended changes. For example, the team reported no problems with the preschool component, praising it for the introduction of preschool labs, the integration of normal and handicapped preschool children, and the exemplary efforts to locate preschool children with special needs. It also commended kindergarten screening but noted one problem: principals and regular kindergarten teachers did not see the value of the screening procedures. The team therefore recommended that these groups be trained in the use of screening instruments and be involved in the actual screening. The full report was distributed to the project director, staff members, and parents.

HUMAN RESOURCE REQUIREMENTS

Self-Study

Staff members are the chief participants in self-study. A steering committee is drawn from the staff to organize and coordinate the evaluation. This group is the keystone of the self-study. It appoints other staff members to various subcommittees, assists the subcommittees in their work, calls and conducts general meetings to examine and approve reports, and prepares the final report. Each subcommittee gathers information on some specific program areas, assesses strengths and weaknesses, suggests changes, and reports its findings to all participants in the study.

After a subcommittee has assembled its results and made recommendations, the entire study staff might meet to discuss the report, ask for clarification, suggest revisions, and approve the report. When all subcommittee reports have been approved, the steering committee assembles the final report and distributes it to the appropriate audience. This may include the program director, the staff, the school board, parents, and other interested members of the community.

For smaller ECT-I programs, or if only a few program areas are to be examined, a hierarchical structure for self-study may be unnecessary. Instead, the whole staff and the director might make up a single committee to identify and study a few key areas. For example, they might decide to assess program goals, chief instructional areas, and staff qualifications. At some other time, they might evaluate other areas such as program administration and facilities.

The continuing involvement of the program staff in the evaluation is the chief strength of self-study. If staff members see a real need for

improvement because of their participation in the study, important changes can result. Self-study can provide a strong link between evaluation results and actual program modification, because those who conduct the evaluation also play central roles in implementing proposed changes.

Outside Review

Outside review procedures also depend on practitioners. As the example cited earlier illustrates, a visiting team may include teachers and administrators from nearby programs, early childhood experts, and state ECT-I staff, who are assigned to subcommittees corresponding to their interests and expertise. For example, an ECT-I program director might concentrate on assessing the administration of the program, while a specialist in early childhood language development would focus on the program's language arts activities.

In planning for an outside review, a staff member or committee is usually designated to coordinate the visit, arrange interviews with the staff, plan tours of facilities, and organize classroom observations. A chairperson is generally selected from among the visiting committee to oversee activities before, during, and after the visit. This person also reports orally to the program staff at the end of the visit and drafts the final report, summarizing the committee's views and recommendations. It is often useful to precede an outside review with a self-study, because the resulting report can help to focus the inquiry.

Like self-study, outside review can be relatively inexpensive, depending upon the extent to which committee members are remunerated, if at all. Moreover, it can provide a fresh and different perspective on the program that internal reviews might miss, and lead to informal exchanges of information as visitors report on their experiences and those of other programs they have visited.

PROCEDURES FOR CONDUCTING SELF-STUDY AND OUTSIDE REVIEW

Self-Study

Self-study data collection depends on the work of several subcommittees, each focusing on a specific part of the program and seeking answers to questions related to that part. For example, the subcommittee examining program structure might raise questions such as: What is the staffing of the program? How is the typical day organized? To what extent and how are individual programs provided for each child?

Identifying questions to guide the self-study, and refining this list, is an important first step. A self-study can easily go awry by spending too much time seeking information to "answer" unimportant questions or examining areas in which a program change is simply not possible. In this regard, it can be helpful to suggest possible findings from each question before conducting the study, and then examine whether these lead in some productive direction. If it turns out that, regardless of how a particular question might be answered, no practical action could result, examining this question in the self-study would be fruitless.

Subcommittees may use several data gathering methods in conducting this work. They can review documents such as program applications, evaluation reports, and curriculum guides. They can interview the program director, principals, and other staff members. They can distribute questionnaires to teachers in the program. While these data collection activities are guided by the list of identified questions, that list may change somewhat as the study progresses.

As information is gathered, it is discussed at committee meetings. The committee members for each self-study area examine the results of their

questionnaires, interviews, and document reviews. They may look for agreement and disagreement among different sources of information. They may also give their own opinions about program strengths and weaknesses. As this informal analysis proceeds, they discuss what they think they have found and work to reach consensus on answers to specific questions. If consensus does not emerge, the alternative points of views then are documented. Finally the chairperson writes a report assessing the current state of the program in the areas examined, recommending improvements and summarizing the findings.

Outside Review

In preparing for a site visit, the visiting committee examines available documents such as grant proposals, program descriptions, curriculum material, budgets, evaluations, and self-study reports. On that basis, committee members identify issues and questions to be addressed. During the visit, these are investigated through interviews and observations. Interviews are usually unstructured and short because of time constraints on the committee and program staff. They might include some brief questioning of teachers following a classroom observation, and more extensive discussions with the program director, principals, and parents. Because committee members cannot participate in day-to-day program activities, a considerable amount of information is drawn from passing observations as they interview staff members, attend meetings, or just roam the halls.

Like self-study, outside review depends on informal data analysis. If a self-study or other evaluations have been done, the visiting committee can consider them and try to build on the findings. Alternatively, the committee might compare the current state of the program, as determined by observations, interviews, and program documents, to some set of criteria for a "good"

early childhood program. These criteria may be drawn from program guidelines, from a state education agency, or from material generated by commercial program developers.

FURTHER CONSIDERATIONS

Self-study requires a great deal of staff time because of the heavy involvement of staff in the actual evaluation process. Unless staff members understand what will be required of them, and how the evaluation will serve to improve the program, they may be reluctant to devote the time and energy necessary to make self-study successful.

The data collection procedures for a self-study need careful attention. It is important to decide in advance what questions to ask about each component of the program. A first step might be to consult a document such as the Elementary School Evaluation Criteria (National Study of School Evaluation, 1973) for guidelines or starting points for self-study, and to modify these as necessary for each specific evaluation. A booklet aimed at elementary schools, for example, may not deal with important early childhood concerns such as screening or parent involvement. At the same time, it will include areas that may be beyond the scope of an ECT-I evaluation, like transportation and support services.

Another self-study issue is the depth and objectivity of data collection and analysis. It is crucial that examination of the program be frank and self-assessment candid. Depending upon who is actually involved in the self-study, the results may be biased in their favor, while shortcomings may be blamed on others, such as administrators, parents, pupils, and the community. Exercising care in planning for the self-study, and in selecting members for each committee, can alleviate these problems.

In outside reviews, because the visits are short, data collection and analysis may not be as thorough as one would like. Thus important program aspects may be missed. Brief sporadic visits to evaluate classrooms have several potential drawbacks. The sample of classroom activities observed may be small and atypical. Teachers and children know that visitors are coming, so that special preparations may have been made. Teachers try to exhibit their best practices, and young children may be distracted by a group of strangers. Moreover, the observers' views of teachers and classrooms differ from children's views; yet the latter may be the most informative for discovering ways to improve the program. Combining brief classroom observations with follow-up teacher interviews can help to identify these problems of perspective. As in the self-study, care should be exercised in the selection of outside reviews in order to obtain a diverse but balanced set of perspectives. (For further discussion of these points, see Scriven, 1977.)

These cautions should not discourage the use of self-study and outside review. The results of such an effort can be quite suggestive, and lead to new programmatic or further evaluation activities. As we shall see, no procedure is without fault, and no method by itself can ensure program improvement. Rather, it seems to us that strong staff interest in conducting the study and good local technical assistance in its design and implementation are critical to successful evaluation and use of the results.

NOTES ON SOURCES OF FURTHER INFORMATION

There are several guides and manuals for conducting self-studies and outside reviews. Among these is material produced by the National Study of School Evaluation (1973), the National Council for Accreditation of Teacher

Education (1970), the National Association of Secondary School Principals (Georgiades, 1978), and the Arkansas Elementary School Council (1971). Of course, these sources can only guide the development of a self-study and outside review protocol for ECT-I programs; each was written for different purposes and hence must be adapted for early childhood programs.

In addition, other sources mentioned throughout this resource book will help direct different aspects of an ECT-I program review. Payne (1951) has written a classic book on framing and asking questions. Ferreira and Burges' (1976) and Boehm and Weinberg's (1977) books on classroom observation are practical and concise. Books by Kerlinger (1973, Chapters 28 and 31), Travers (1978, Chapters 8 and 12), and Tuckman (1972, Chapters 7 and 18) have sections on interviews and observation, which should be helpful to anyone planning and carrying out self-studies and outside reviews.

III. QUANTITATIVE METHODS

We have grouped together in this section three evaluation methods based on quantitative social science research techniques:

- Structured process evaluation, which focuses on how services are delivered
- Product evaluation or examination of program outcome data, which considers the impact of services on specific program objectives
- Process-product evaluation, which combines aspects of the other two to analyze the link between program practice and program outcomes.

STRUCTURED PROCESS EVALUATION

Structured process evaluation examines aspects of how ECT-I services are actually delivered and how well this fits with what the program was designed to do. Such an evaluation might, for example, describe general features of a program, such as the number and kinds of students served, what materials are available, and how classrooms are arranged. It can also provide some fine-grained information about the frequency and duration of specific activities and behavior, such as the incidence of children's aggressive acts or the time teachers spend in direct reading instruction. When evaluators and ECT-I staff compare these data on classroom practices with the program objectives, they may uncover some areas of discrepancy. Such information can be helpful in identifying aspects of program practices where attempts at change are warranted.

PRODUCT EVALUATION

Product evaluation builds naturally on a short-term program impact evaluation that might be conducted for purposes of accountability (see the resource book Short-Term Impact Evaluations of Early Childhood Title I Programs by Haney, 1980). Here, however, the focus is on a more detailed

analysis of where and with whom the program succeeds or does not succeed. Such an evaluation might examine, for example, the degree to which certain program objectives are achieved or the variability in program effectiveness across classrooms, schools, or subpopulations of children. If adequate data exist on program outcomes, careful examination can indicate the need for improvement with respect to specific objectives, program sites, or groups of children. In addition, product evaluation can provide guidance on where to focus efforts for change.

PROCESS-PRODUCT EVALUATION

Process-product evaluation combines elements of both of these methods to investigate the links between specific program practices (e.g., the amount of time children spend in reading instruction) and child outcomes such as norm- or criterion-referenced test scores. This method, by examining program practices jointly with their consequences, can suggest direct ways of improving the program. Of the three techniques discussed in this chapter, process-product evaluation offers perhaps the greatest potential for program improvement. It is, however, the most complicated and difficult to implement successfully.

EXAMPLES OF QUANTITATIVE METHODS

A Structured Process Evaluation*

A short-term impact evaluation of an early childhood program indicated that students were not performing as well as expected. The local director of evaluation, drawing on recent research by Berliner and Rosenshine (1976), suggested to the program staff that these results might be related to the

* This case was adapted from several sources: Issac (1977), Fenstermacher (1977), and Behnke, Bennett, Chase, Day, Lazar, and Mittleholtz (no date).

time students actually spent on learning basic skills. After a series of discussions among the evaluation and program staffs, together they decided to focus further evaluation on the frequency of classroom disruptions, their apparent causes, and how teachers tried to deal with them.

The evaluation staff assembled eight observers in four teams of two. For six weeks these teams observed eight classrooms, using a checklist developed by Berliner and Rosenshine to code the type and number of disruptions and teachers' ways of coping with them. The observers also noted how effective each coping technique was in terms of how much instructional time was lost, whether teaching was interrupted, and how long the disruption continued after the technique was used.

The evaluators found that most disruptions were caused by the students themselves, that they usually occurred during reading instruction, and that a few children were mainly responsible. Nonverbal techniques, such as isolating the students who were causing the disruptions, were found to be particularly effective in reducing the loss in instruction time.

A Product Analysis*

The coordinator of a pilot preschool program approached the field research center of a nearby college for advice on how to evaluate their pilot effort. A research associate at the center persuaded him to develop a criterion-referenced test linked to the objectives of the program. Such a test, she suggested, could provide three kinds of information: individual child data to help in monitoring each child's performance, classroom data to assist teachers in focusing efforts on areas needing improvement, and

* This example is fictional, although the original idea came from Airasian and Madaus (1972), and the format of the tables from Morris and Fitz-Gibbon (1978, pp. 136-143).

more general program data to enable the director and staff to identify weak program components. After several meetings to clarify program objectives, a criterion-referenced test was produced.

The results from this testing produced considerable information. Table 1 shows some of the data that could be provided to the classroom teacher. From this, the teacher can see, for example, that the first child has done well on test items measuring all four objectives and should move on to other work, while Larry, the second child, appears to need more instruction in all areas. Table 2 shows results from the testing summarized by classroom. In this display, teachers can see overall strengths and weaknesses in their classes, and so can behave accordingly. For example, the teacher in Classroom 1 can easily see that more effort should focus on Objective 4, capital letter recognition; and that, while certain children need help in learning to write their names, no overall class time need be devoted to this objective. Finally, Table 3 provides summary information on the whole program. From this, the program director can readily see that Objectives 3 and 4 are not being met, and this suggests a further examination of instructional activities in this area.

A Process-Product Evaluation

Some of the cost of day care in this country is subsidized by the federal government. In order to assure that adequate care is provided, centers that receive such subsidies must meet certain standards of physical plant and human resources. Of particular concern were the regulations governing staff-child ratio, staff training, and staff experience. To help determine the appropriateness of these standards, the Administration for Children, Youth, and Families (ACYF), the federal agency that helps to

Table 1

Results from One ECT-I Classroom Showing Which Objectives Each Child Achieved

	Objective 1 Writes Name	Objective 2 Knows Colors	Objective 3 Reads Lower-Case Letters	Objective 4 Reads Capitals
ECT-1 Classroom 1				
Nancy	+	+	+	+
Larry	-	-	-	-
David	+	+	-	-
Christine	+	-	-	-
Charles	+	+	+	-

Table 2

Percertage of Students Achieving Each Objective

ECT-I Classroom	Objective 1	Objective 2	Objective 3	Objective 4
1	80%	60%	40%	20%
2	99%	80%	50%	30%
3	98%	70%	35%	15%
4	90%	75%	45%	25%
5	85%	65%	35%	20%

Table 3

Overall Percentage of Students Achieving Each Objective

Objective 1	Objective 2	Objective 3	Objective 4
90%	70%	41%	22%

develop the regulations, decided to explore two issues: (1) how these standards influence the behavior of workers in the day care centers, e.g., how much time is spent teaching children; and (2) how this in turn affects children's behavior and development. ACYF contracted with Abt Associates to investigate these questions.

Developing procedures for collecting data on each of these elements was the first major task of the Abt staff. To obtain background data on day care workers, the staff in a sample of centers were asked to complete a questionnaire. Other data, such as staff-child ratio, were collected during several observations made in the same centers. Information on the behavior of caregivers and children was gathered using structured observation protocols. Finally, to assess the children's academic progress, the research team administered the Caldwell Preschool Inventory and the Peabody Picture Vocabulary tests, first in the fall and then again the following spring.

These procedures generated a large amount of data, which were analyzed with sophisticated statistical techniques. The analysis identified group size--the number of children present--as an important feature of the day care center. Group size was strongly related to how caregivers spent their time: for example, in smaller centers caregivers spent considerably more time interacting with children than in larger centers. Amount of interaction, in turn, was strongly associated with positive child experiences (e.g., more time engaged in mastery tasks) and better academic development (i.e. larger gains on the standardized tests). These results suggested to ACYF that future day care center regulations should give more attention to group size.

Common Features

Each of the examples discussed above represents a highly focused form of inquiry. Each research process began as a result of some perceived problem. In the first example of structured process evaluation, for instance, concern arose about the program's effectiveness. Out of this, some fairly specific research questions emerged. Discussion between program staff and evaluators, in that example, led to the decision to focus on classroom disruptions, their apparent causes, and teachers' attempts to deal with them. An appropriate evaluation plan could then be developed.

A key feature in developing the evaluation plan, as seen in each of the cases described above, is the selection, or if necessary the creation, of suitable data collection procedures for each area of study. In the product analysis example, this involved the creation of a criterion-referenced test (CRT) that was geared to the objectives of the new curriculum. In general, the quantitative methods described in this section all rely on standardized instruments--especially structured interviews, structured observations, and standardized tests.

Conducting these evaluations often requires time from program personnel and extensive support from evaluation staff. This collaboration is needed at each step of such a study--beginning with the identification of concerns and development of the evaluation plan, through the collection and organization of data, and concluding with data analyses and report writing. In principle, quantitative methods can be very efficient evaluative tools when the program manager or policy maker wishes to undertake focused inquiry into issues that have been fairly well specified in advance.

HUMAN RESOURCE REQUIREMENTS

Structured Process Evaluation

Structured process evaluation usually requires the participation of a program evaluator or evaluation team. In our example, several persons were involved. They created or selected appropriate evaluation instruments, trained interviewers and observers, directed the evaluation, and wrote the reports. Often, the evaluator may participate directly in data collection by observing classrooms, interviewing teachers, aides, and parents, and distributing questionnaires. In small LEAs with limited resources, a part-time person can carry out the entire evaluation, from data collection to report writing.

In general, structured process evaluation depends on a knowledge of survey research methods for developing an efficient plan for data collection, devising valid interview guides and observation protocols, and carrying out the data collection. The quality of this kind of study in large measure depends upon the adequacy of the data collection instruments. In some LEAs, local evaluation staff in collaboration with program staff have the capacity to develop such instruments. Elsewhere, help can often be obtained from outside sources.

Some care also needs to be exercised in selecting the staff who will collect the data. Interviewers should be able to put respondents at ease so as to encourage free and full discussion. Moreover, they must be able to record responses without showing surprise, approval, or disagreement. Observers of young children in ECT-I classrooms should "blend into the woodwork" to avoid distracting those they observe. They must be skilled listeners and watchers, accurately recording what they see and hear. (See Burges, 1976, for further discussion.)

Product Evaluation

The data in a product evaluation usually consist of students' scores on norm- or criterion-referenced tests. Teachers and aides are frequently called on to collect this information. Unfortunately test administration is sometimes haphazard, thereby limiting the validity of the data and the usefulness of the evaluation. Assuring proper test conditions is thus an important consideration.

The development of CRTs requires specially trained personnel. Sophisticated psychometric methods are needed to specify knowledge domains, formulate questions, ensure test reliability, and determine test length. In most LEAs, development of local CRTs will require outside help. This is not needed if commercial CRTs are used or if less sophisticated, objective-referenced tests are created.

Human resource requirements for analyzing test results can vary considerably. On the one hand, few local resources are required if the LEA buys an analysis service from test companies. If commercial test analysis is inadequate or if the test is a locally developed instrument, more sophisticated LEA skills will be required. Program staff will need to collaborate with an evaluator or other person with psychometric training in order to translate test results into suggestions for program modification.

Process-Product Evaluation

Process-product evaluation draws on both structured process data and product information. (See the discussion above for data collection considerations.) In addition, because analysis requires complex statistical methods to relate process and product data, expertise in statistics and computer-based data analysis is needed. In LEAs with an evaluation office,

these skills are often available. Other LEAs may need to seek outside help.

PROCEDURES FOR CONDUCTING QUANTITATIVE EVALUATIONS

Procedures for conducting quantitative evaluation by any of the three methods described break down into four categories:

- Determining the focus of the evaluation
- Choosing or creating data collection instruments
- Collecting the data
- Analyzing the results.

For both structured process evaluation and product evaluation, the first three activities are the most important; if good data are collected, data analysis often will be relatively straighforward. Process-product evaluation, by contrast, not only depends on the quality of the data but also requires their extensive analysis to produce useful information for program improvement.

Structured Process Evaluation

This type of evaluation usually begins with a question, hunch, or theory about how the program can be improved. In our example, student outcomes indicated that the early childhood program was not as effective as had been hoped. One evaluator's experience with time-on-task studies led him to suggest a study on the use of classroom time. Framing evaluation questions in advance is necessary because the data collection instruments--whether they are questionnaires, interview forms, or observation protocols--have fixed questions and responses. Thus, there is little opportunity to redirect the inquiry in midstream if new concerns arise.

Checklists and coded behavior records are common data collection procedures. Observers might use a checklist to count occurrences of certain classroom behavior, such as the number of times a child's attention wanders during a 15-minute group activity. For more detailed information, coded behavior records are often used to count all acts that occur within some fixed time interval. This observation procedure requires the development of a set of general behavior categories, and a set of rules for assigning specific behavior to one such category. The definition of categories depends upon some preconceived notions about what is important to look at. The Flanders teacher-student interaction instrument typifies this approach. As Figure 1 shows, its focus on the verbal interaction of teacher and child is reflected in the structure of its behavior categories.* Other observation systems might focus on social interactions among peers and adults, or on the use of materials, or space, or some combination of these.

Structured interviews represent another common data collection procedure. Again, the program staff together with the evaluator must decide what information they want to gather. The program theory, program plans and activity lists will suggest what information may be needed. For each of these areas, specific questions are developed. Next, the interview form is created, field tested, and revised where necessary. Finally, the interview instrument is ready for administration (Burry, 1978).

If the questions for structured process evaluation are well thought out and if data collection is done carefully, data analysis should be straightforward. A simple presentation of the results of interviews or

---

* See Flanders (1965) for further discussion.

Teacher Talk	Indirect Influence	1. Accepts feeling: accepts and clarifies feeling tone of the students in a non-threatening manner. Feelings may be positive or negative. Predicting and recalling feelings are included. 2. Praises or encourages: praises or encourages student action or behavior. Jokes that release tension, not at the expense of another individual, nodding head or saying "uhhuh?" or "go on" are included. 3. Accepts or uses ideas of student: clarifying, building, or developing ideas or suggestions by a student. As teacher brings more of his own ideas into play, shift to category 5. 4. Asks questions: asking a question about content or procedure with the intent that a student answer.
	Direct Influence	5. Lectures: giving facts or opinions about content or procedure; expressing his own ideas; asking rhetorical questions. 6. Gives directions: directions, commands, or orders with which a student is expected to comply. 7. Criticizes or justifies authority: statements intended to change student behavior from non-acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing, extreme self-reference.
Student Talk		8. Student talk-response: talk by students in response to teacher. Teacher initiates the contact or solicits student statement. 9. Student talk-initiation: talk by students, which they initiate. If "calling on" student is only to indicate who may talk next, observer must decide whether student wanted to talk. If he did, use this category.
		10. Silence or confusion: pauses, short periods of silence, and periods of confusion in which communication cannot be understood by the observer.

(Source: Tuckman, 1972)

Figure 1: Summary of Categories for Interaction Analysis

observations can be instrumental in identifying areas for program improvement. For example, one study of the use of class time used pie charts to display the amount of instructional versus non-instructional time. The surprise and consternation about how little time was spent on teaching and learning led to efforts that increased the time devoted to teaching and reduced other activities.*

Product Evaluation

Care in selecting or developing an appropriate test is essential to the success of product evaluation. The usefulness of the data depends on a precise explication of program concerns and subsequent research questions, choosing or creating a test to match these needs, and careful data collection. Commercial tests may not be targeted at areas of local interest; early childhood tests, for example, often provide little help for assessing the socio-emotional goals of a program. Thus, ECT-I staff should carefully examine the content of any test to ensure that it measures what they want to evaluate. If commercial tests are unsatisfactory, the staff may consider creating a local instrument; but this can be expensive and time-consuming.

Similarly, exercising care in test administration is particularly important with young children. The resource book Assessment in Early Childhood Education by Haney and Gelberg (1980) discusses this question in some detail. In addition, the Standards for Evaluation and Psychological Tests (APA, AERA, & NCME, 1974) is a good guide to appropriate practices in test administration.

If these suggestions about test selection and administration are heeded, analysis of the test data should be quite straightforward and informative.

* For a discussion of this evaluation, see Holley (1978).

As we saw in our example of product evaluation, simple tables of test results can dramatically reveal both program strengths and areas needing improvement.

Process-Product Evaluation

Process-product evaluation, as we said earlier, combines structured process data with the examination of outcome measures to find ways to improve the program. Our comments on conducting structured process evaluations and product analyses apply equally to process-product studies. Evaluation questions must be framed in advance; data collection instruments must be chosen or created that are congruent with the evaluation issues; and process and product data need to be carefully collected.

Data analysis, however, plays a larger part in process-product evaluation than in either of the other two quantitative methods. Identifying the relationship between process data from questionnaires, interviews, and observations on the one hand and from test scores on the other requires sophisticated statistical techniques. Further, ECT-I personnel need to interpret and use the results with caution. Statistical analysis may indicate that some set of program variables is associated with higher test scores, but this does not necessarily mean that the higher scores are due to these factors. For example, a process-product evaluation might show that children who attend full-day kindergarten do worse in the first grade than those attending only half days. But before concluding that the full-day program should be modified, other hypotheses should be examined. For instance, are children who are expected to do poorly in school selected for the full-day sessions? If so, it may not make sense to compare the test scores of children across these two programs. To draw valid inferences here requires

both a knowledge of the local program and its context, and statistical expertise. Smaller LEAs can consult a statistician or outside evaluator to assist in such an effort.

FURTHER CONSIDERATIONS

In deciding to use one of these methods, LEA personnel should consider several issues. In general, program staff need a clear idea of the questions they want to investigate, since the structured data collection procedures required here must be created or selected before the research begins. In our example of a structured process evaluation, for instance, the evaluators wanted to examine classroom disruptions; therefore, they found instruments that coincided with this intention. ECT-I staff should also realize that not all aspects of the program can be examined with quantitative methods. For instance, since interviews and observations are usually brief and take place during one or a few sessions, these techniques are not well suited for investigating program interactions in depth. The qualitative techniques discussed in the next chapter are more responsive in this respect.

NOTES ON SOURCES OF FURTHER INFORMATION

Structured Process Evaluation

Many authors discuss evaluation methods that are similar to structured process evaluation, although they often call their approaches by different names and have somewhat different purposes in mind. Some treat formative evaluation (for example, Baker, 1974; Scriven, 1966; and Weiss, 1972). Others talk about examining program implementation (see Hall & Loucks, 1977; Morris & Fitz-Gibbon, 1978; Rossi, Freeman, & Wright, 1979; and Wolf, 1979). All of these sources contain potentially helpful advice for carrying out a structured process evaluation.

In addition, many sources discuss the data collection techniques used in a structured process study. Maccoby and Maccoby (1954), Cannell and Kahn (1953), and Sellitz, Wrightman, and Cook (1976, Chapter 9) contain good discussions of interviewing techniques. Medley and Mitzel (1963), Weick (1968), Rosenshine and Furst (1973), Cassel (1978), and Stallings (1977) outline observational strategies that are applicable to structured process evaluation.

Product Evaluation

Innumerable books and articles discuss tests and measures. Here we list some that are most relevant to interpreting product evaluations. Several resource books in this series on ECT-I evaluation present useful discussions. Haney and Gelberg (1980) examine issues in the assessment of young children. Haney (1980) outlines short-term impact strategies. Kennedy (1980) looks at longitudinal data as a means for evaluating ECT-I programs. Other works on tests and measures in an evaluation context are Sax (1974), Millman (1974), Airasian and Madaus (1972), and Morris and Fitz-Gibbon (1978).

Process-Product Evaluation

In addition to the sources mentioned, Walberg (1974), Cooley (1978), and Pedhazur (1975) discuss practical and analytical issues related to process-product research. The National Day Care Study (Roupp, Travers, Glantz, & Coelen, 1979) is a good example of this approach.

IV. QUALITATIVE METHODS

This section describes three qualitative inquiry methods:

- Documentation
- Investigative evaluation
- Ethnography.

These methods resemble each other (and differ from others) in that they are based on the idea of "the evaluator as instrument." Less emphasis is placed on standardizing data collection activities, and much is left to the individual evaluator to decide as the work proceeds. Because so much is left to the evaluator's discretion, the choice of person or persons to conduct the study is critical.

Qualitative methods are very flexible, powerful research techniques capable of in-depth, broad-based description and analysis. Interpretation of early data can substantially influence subsequent research activities. Although the three methods are similar in some ways, they differ in the time they require, documentation taking the least time and ethnography the most. They also emphasize different data collection methods. Documentation depends mostly on record examination; investigative evaluation on intensive interviews; and ethnography on observation.

DOCUMENTATION

Documentation aims at the accurate depiction of ECT-I programs. The documentors attempt to collect and examine evidence in order to portray the program clearly and truthfully. Documents may include funding applications, teachers' plans and self-reports, interviews, questionnaires, notes from observations, examples of children's work, photographs, and diagrams of classrooms.

Documentation can portray the program in depth by describing representative pieces of it. It can also assess whether the program has been implemented as planned by juxtaposing planning documents and testimony about what the program should be, and data on what the program actually is.

INVESTIGATIVE EVALUATION

This method uses intensive interviewing, observation, and document examination. We call this investigative evaluation because it resembles detective work or investigative reporting. The evaluator forms hunches, follows leads, verifies sources, and throws out hunches that can't be supported. This sort of evaluation can address a variety of questions. In examining program operations, it might assess what the program does, how it works, what its goals are, and whether these goals have changed. In looking at implementation, the evaluator might focus on the congruity between goals and program reality, and between legal mandates and program practice. In suggesting program improvements, investigative evaluation can help determine what went wrong and why, and offer recommendations for change.

ETHNOGRAPHIC EVALUATION

Ethnographic evaluation requires long-term in-depth observation of ECT-I programs. Through such extensive observations, the ethnographer attempts to understand the program and the meaning of the events to the various participants such as parents, children, and staff. In the course of these observations, the ethnographer might be interested in issues such as how the program operates in different school situations, what those involved with the program identify as its strengths and weaknesses, and how the program influences young children's school experiences. Such an

evaluation can provide a rich description and portrayal of what it is like to participate in the program. Finally, ethnographic evaluation attempts to discern the chief features and processes of programs and to provide a comprehensive view of their operation.

EXAMPLES OF QUALITATIVE METHODS

Documenting a Marine Science Program

Documentation was used to examine a program in marine science for children in kindergarten through fifth grade. This cooperative project between an aquarium and a magnet elementary school was intended to increase vocabulary, improve observation and reasoning skills, stimulate art work, and build interest in science. Aquarium staff, the program director, and teachers wanted to know how well the program was doing and how activities and processes could be improved.

By talking with staff, teachers, parents, and administrators the documentor determined that program administrators were the main audience for the evaluation, but that the aquarium staff, teachers, parents, and state and local officials were also interested parties. In their first meeting, the documentor and the project director framed evaluation questions based on program objectives. Program staff wanted evidence of such things as increased vocabulary, improved observation, and interest in science. They were also concerned about improving discipline at the aquarium. Data gathering procedures were then matched with each question. For example, interest in science was assessed by interviewing children and their parents. Observing at the aquarium and interviewing teachers provided data about children's behavior. Also at the first meeting, agreements were made about who would collect what data. The documentor offered to make observations

and conduct interviews; teachers volunteered to keep diaries; parents and aides agreed to take photographs and to collect examples of children's work.

During the evaluation, the documentor tried to get a wide and representative view of the program. She talked with several groups, including teachers, administrators, parents, students, and aquarium staff; sought out people with different viewpoints on the project; and made sure that all perspectives were included in the final report.

The documentor communicated her findings in two ways. First, after observing activities and interviewing staff members, she discussed her reactions with them. Her comments were as descriptive and nonjudgmental as possible. For example, she told several teachers that, when they left their rooms during class to have coffee, instruction and discipline broke down. On the basis of this information, school staff worked out a plan for some teachers to supervise classes while others took their breaks.

Second, at the end of the evaluation she produced a written report that was distributed to all interested parties. This final report was descriptive, but here judgments were more explicit. One finding indicated that although the program encouraged observation and reasoning both at the aquarium and at school, some children were not as involved as one would hope. The documentor substantiated this with descriptions from her observations, excerpts from interviews, and photographs. She recommended that teachers, not aquarium staff, take charge of maintaining order during field trips.

Investigative Evaluation of a Bilingual Program*

This evaluation focused on implementation of a California bilingual education law that mandates programs for limited-English-speaking (LES) and non-English-speaking (NES) pupils. Since new bilingual legislation was being considered, the evaluators first determined what issues concerned their chief client, the legislature. In consultation with individual legislators, a set of key questions was identified: How are children selected for bilingual programs, are state mandates for bilingual classes --such as prohibiting segregation of NES children--being followed, and are the required annual evaluations useful?

The evaluators visited 17 school districts, and altogether 19 schools. Districts were so chosen as to ensure variability in grade level, location, and bilingual concentration. The evaluators, using primarily unstructured interviews, spent a few days talking with key people in bilingual programs in each LEA. These included parents, regular and bilingual teachers, the bilingual-program director, and the superintendent. The evaluators collected the following information: how bilingual pupils are identified and assessed, how program staffs determine when pupils no longer need bilingual education, how many years pupils participate in programs, what other compensatory programs NES and LES children participate in, how LEAs evaluate their programs, and how much pupil identification, testing, and evaluation cost.

* Because investigative evaluation is relatively new, there are few examples of local use. This example draws from a statewide evaluation meant for legislators and their aides (Bissel, Christophel, Sequeira, and Farias, 1979). It does illustrate, however, the kinds of information such an evaluation can produce, and most of the procedures could be used by LEAs.

The investigators found several problems with bilingual programs in the districts they visited. First, state guidelines for bilingual programs were unclear to many local staff members. Second, districts were incurring unnecessary costs because of inefficient or unnecessary testing and assessment. Third, the responsibility for bilingual education was too diffuse within the state department of education. Fourth, there were inherent difficulties--such as limitations in the state of the art of bilingual testing--in meeting some of the statutory requirements. To alleviate these problems, the report recommended that the department of education consolidate the administration of bilingual education in the state, and identify for the legislature those requirements of the law that were difficult or impossible to apply.

An Ethnography of a K-to-5 Computer-Assisted Instruction (CAI) Program

When the staff of a regional educational agency began a computer-assisted learning program for kindergarten and elementary school children in several rural mountain towns, they asked two ethnographers to evaluate the project. Although they set out to assess the outcomes of the program, the evaluators soon realized that they had to refocus the evaluation. Their central goal became investigating the difficulties that arose when sophisticated educational technology was brought into a rural, economically deprived region of the country.

After initial visits to program sites and discussions with project participants, the ethnographers formulated four areas to investigate: the impact of CAI on teachers' decision making, on teacher-student relations, on classroom social structure, and on children's behavior and personality. As the evaluators later admitted, these four areas only roughly guided the actual evaluation.

During the first year of the project, the evaluators spent 94 person-days in the field in intermittent two- and three-day visits. They kept three kinds of records: on-the-spot observations of such events as children working at the computer terminals; summaries of observations, informal interviews, and discussions; and documents such as computer printouts. Altogether, they collected 900 pages of raw data, which they eventually used to describe and evaluate the program.

A brief summary of the results loses the flavor and fullness of their evaluation; we therefore present a few excerpts instead. The evaluators grouped their findings under three broad categories: technical difficulties, teacher use of CAI, and pupil behavior at the terminals. Regarding the first, they concluded that the system was never fully operational, leading to continuing teacher and pupil frustration throughout the year. They related vignettes from their field notes to substantiate this conclusion. Perhaps the most poignant episode concerned the girl whose printout read "Cry again" instead of "Try again." Her response to her teacher was, "Oh, Mrs. Martin, I could just cry and cry again" (p. 7).

Teacher use of the system varied considerably. For example, the evaluators noticed wide differences in the integration of CAI with classroom instruction. One excerpt from their field notes illustrates the lack of integration:

> I made a specific point of checking with two of the [first-grade] girls about where they were currently in their math lessons. I wondered how closely the drills that they were taking corresponded to what they were doing in class. Apparently it is not very close. Ruth told me that the day's classroom lesson was on "writing mathematical sentences." The drills, however, were all simple addition and subtraction problems. (p. 20)

The findings on pupils' behavior at the terminals are the most intriguing aspects of the evaluation. In general, the evaluators discovered complex interactions between children and computer and among children working at the terminals. For example, they found that CAI generated unanticipated competition in some classrooms.

> Midway through the morning I happened to notice three boys working on the terminals. They made an effort to start together, and it was a real contest. It should be noted that the three boys were not on the same lesson. Nevertheless, there was a great deal of competition to see (1) who would finish first, and (2) who would get the highest percentage. The boy who finally did finish first raised his arms above his head like a boxer and crowed rather exaltedly, "I won, I won." The sweet smell of success was even greater when he found out that he had achieved a higher percentage score on his test than either of his two buddies. Both of them looked a little bit crestfallen, particularly the boy who ended up last. (p. 31)

These excerpts illustrate the kind of description characteristic of ethnography, which can lead the reader to a deeper understanding of what the program in operation was like. We encourage anyone with further interest in ethnographic techniques to read the whole evaluation report (Smith and Pohland, 1974).

PROCEDURES FOR CONDUCTING QUALITATIVE EVALUATIONS

Human Resources

Documentation. The documentation of the marine science program illustrates how several groups can be drawn into the evaluation process. Teachers might take photographs of the classroom and special activities, keep folders of children's work, and record anecdotes. Parents might keep time-activity charts, write up classroom notes, summarize questionnaires, and take pictures. Aides might keep folders on children's work up to date, make diagrams of the classroom, and collect statistical information. Administrators can provide

an overview of the program and information about its environment and context. The outside documentor organizes the documentation, creates information gathering systems, observes the program, analyzes data, and makes recommendations for improvement.

Involving teachers, aides, parents, and administrators as well as a trained evaluator is a chief strength of documentation. Because several groups generate and collect documents, diversified views on the program are provided. The use of program staff and parents to collect data can also help to reduce evaluation costs.

The documentor is, of course, the key person in documentation. Since documentation is not one procedure but several, documentors must tailor the techniques to the particular situation. Different data must be collected to assess different goals and to answer different questions. As the marine science documentation illustrated, the documentor must be familiar with a wide variety of evaluation tools and know which are best for answering a particular question.

To document a program fairly and accurately requires a good deal of sensitivity and judgment. The documentor of the marine science program, for example, took pains to make sure that all opinions about the program were represented in the final report. Only a skilled documentor can keep the process from threatening those it is meant to enlighten. Certainly, keeping the evaluation factual rather than judgmental, at least in the initial stages, helps to promote this goal. At the same time, facts themselves can be threatening, especially when they inform people that what is happening is not what they want to happen. If care is not exercised, the focus of the evaluation can easily shift from learning about a program to

fault-finding and excuse-making. Further, the links between the documents collected and program problems are not always clear; nor are strategies for program improvement always self-evident from the documentation report. Thus the documentor's knowledge about the program and experience in educational contexts are important assets for teasing out relationships and suggesting improvements.

Investigative evaluation. Unlike structured interviews and observations, the intensive interviewing involved in investigative evaluation is much like the work of a journalist or a criminal investigator. Good intensive interviewing depends not only on understanding interview techniques but on adaptability to contingencies, good guesswork, and intuition. The quality of an investigative evaluation thus depends on the skepticism, instinct, integrity, intelligence, and to some extent luck of the evaluator.

Ethnographic evaluation. As one can infer from the ethnographic study of the CAI program, ethnographic evaluations require experienced observers. Observers can be either nonparticipants, passively observing children and staff, or participants, actually becoming part of what they observe. Both evaluators of the computer-assisted instruction program were nonparticipant observers. In another evaluation, Smith and Geoffrey (1968) used both participant and nonparticipant techniques to study an urban classroom. Geoffrey, the classroom teacher and participant observer, recorded at the end of each day his perceptions of what went on in class. Smith, the nonparticipant, observed the class and kept detailed notes on what he saw. Thus the evaluators obtained both inside and outside views of what took place and why.

Schatzman and Strauss (1973) provide a good impressionistic description of an ethnographer:

> [He] is a learner, has patience, is tolerant and sympathetic. He wonders first and judges last; he appears to be that way and is that way. Furthermore, he generally accepts whatever he sees and hears at face value; he denigrates no motives. He does not visibly take sides on arguments . . . no matter how much he may be invited to do so. He is open to the discovery of whatever is not so obvious to others. He is most considerate, polite, but not shy; he is, in fact, rather tough in the sense that he cannot be put off for too long, nor shamed or coerced. He cannot be bought off or drawn into private arrangements, even to gain the data he needs. He assumes that the hosts ultimately would have it no other way. (p. 65)

Qualitative Data Collection and Analysis

Documentation, investigative evaluation, and ethnographic evaluation depend primarily on document examination, unstructured interviews, and direct observations. Although all three methods may employ all these procedures to some degree, record examination is the chief procedure for documentation, intensive interviewing for investigative evaluation, and long-term observation for ethnographic study.

Documentation: Examining records. The documentor can obtain information by systematically collecting existing records and by creating procedures to gather additional information not currently available. It is often useful to begin by simply listing the most important activities of the program, and the records that are normally maintained by program staff. When existing records provide insufficient information about particular activities, new procedures must be devised to gather such data. Figure 2 shows a chart of activities and records from a first-grade documentation; for each activity of the program, the documentor identified existing records or suggested new record-keeping practices where necessary.

There are several ways to analyze and present results from classroom documentation. Observers can summarize and transform their notes and checklists into a narrative of what occurred in the classroom while they were present. Such a narrative helps describe how the program appeared to the observer:

> 9:00 Skills Time: Children arrive, sign up for lunch, check the mini-lab board to choose an activity for the end of the day, and gather in two groups for a short meeting to plan the day.
>
> The 5 and 6 year olds choose from a variety of activities, and meet in small groups or individually with the teachers to work on language experience activities, reading, or number concepts.
>
> The 7-9 year olds work in their individual study carrels with partners at reading, writing, math or current projects. They confer individually with teachers to plan their individual goals or projects. Parents help individuals or small groups. Another teacher or aide gathers small groups for skills sequence teaching or listens to children read from their self-chosen books.
>
> At 10:30 the groups begin to flow from room to room as the children work on self-chosen activities at the centers in the school. They may cook, do a play, work at the science center, listen to recorded stories, paint, play games, or continue on "school work." Skills centers are assigned weekly to the 7-9 year olds in the areas of writing, math, and reading. The groups gather again for a story and discussion period at the end of the morning, and to hear their current story be read aloud.
>
> (Engel, 1975, p. 61)

Categorizing observations can also be useful for analysis and presentation. For example, a visitor can observe a teacher for one hour, noting all his activities and categorizing them according to purpose. Data of this kind illuminate a variety of questions such as how teachers spend their time. In addition, graphic displays can assist in analysis. For example, the evaluator can observe children's choice of activity for a set time, determine the frequencies of choice, and graph them. Graphs can help

Activities (First Grade)	Record or Document to be Used for Monitoring the Activity, and Assessment of its Adequacy	Regularity of Document Collection and Assessment of its Adequacy
Reading Aloud with Teacher or Aide (3 times per week)	Teacher/aide's record book: gives dates of reading and number of pages read ADEQUATE	Constant documentation ADEQUATE
OR		
Reading Aloud into a Tape Recorder at Recorder Corner (3 times per week)	Aide's recording form: gives amount of time, progress, distractions ADEQUATE	Only on discipline problems INADEQUATE (Ask staff to keep records on all students)
Reading Seatwork: Choice of		
- Workbook	Completed workbooks: record of pace and completion ADEQUATE	Constant documentation ADEQUATE
- Library book reading	Library book use INADEQUATE	No documentation INADEQUATE (Have students record library books read)
Perceptual-Motor Time, e.g. 15 minutes per day in school gym with aide	No records INADEQUATE (Have aide keep a checklist of length and content of daily sessions)	
- Clapping rhythm exercise (in group)	None INADEOUATE (Ask aide to keep notes on this activity)	
- Open balance (individual, on balance beam, jungle gym, etc.)	None INADEQUATE (Ask aide to keep notes on this activity)	

(Adapted from Burry, 1978, pp. 12-13)

Figure 2: Program Activities and Records

depict the popularity of classroom activities for each child and for the whole class.

Investigative evaluation: Intensive interviewing. The central procedure for investigative evaluation--intensive interviewing--is most useful when neither the interviewer's exact questions nor the respondent's likely answers can be specified in advance. Collecting data by this procedure seems particularly well suited for discovering the perceptions, attitudes, and motivations that, when sought explicitly, respondents may be unable or unwilling to provide.

The investigator begins the evaluation by "scouting the scene"--reading all available program documents, and doing preliminary interviews, often by phone. This background work helps the investigator to become familiar with the program, to identify key individuals, informants, and sites, to begin formulating key questions, and to develop hunches and initial perspectives on program problems.

The investigator is then ready to begin intensive interviews with key people. Intensive interviewing is quite different from the structured, standardized interviews we have already discussed. Rather than asking everyone the same questions, the interviewer tailors questions to each person and situation. What questions are asked and how they are phrased depends on the evaluation issues, the circumstances of the interview, and the respondent's reaction to previous questions. Intensive interviews are aimed both at collecting valid, factual information, and at gathering remarks and anecdotes that reflect views of the program in action.

Data collection and analysis in investigative evaluation are closely linked. These two activites go on almost simultaneously: preliminary

analysis indicates additional data to collect; as further data are collected the evaluation's hypotheses can be further examined. In short, data analysis goes on throughout the evaluation, even during report writing. While collecting data, the investigator forms and checks out hunches about program facts, frequencies, and causations. He corroborates facts by looking for confirming and contradictory evidence. If he finds substantial evidence consistent with the hunch, his confidence increases; contrary evidence leads him to discard the hunch.

Ethnography: Long-term observation. In some ways, long-term observation resembles intensive interviewing. The evaluator maintains considerable flexibility in determining what is to be observed, and the technique is particularly useful when complex behavior is to be studied. As with intensive interviewing, background work is a prerequisite of data collection. Similarly, the pieces of information collected through observation are not automatically accepted as representative or valid; rather, the ethnographer attempts to verify what he sees by observing many situations, just as the intensive interviewer checks out information obtained from each respondent.

The initial step in ethnographic data collection is deciding what to observe. Just as the investigative evaluator scouts the scene, the ethnographer often begins with some unfocused observations. He views as many aspects of the program as possible and "over-records" his observations. As time passes, he begins to identify representative times, people, and situations, and to form hunches and hypotheses about what happens and why. Ideally, there is a pattern to this process: the initial focus on very general observations spirals inward to increasingly specific observations of key events, relationships, and the like. The ethnographic evaluation of the CAI program illustrates

this. The evaluators initially focused on broad topics such as teacher decision making and classroom social structure. By the time they had finished the evaluation, they reported on more specific topics such as technial difficulties in implementing the program and children's behavior at the computer terminals.

During both general and focused observations, field notes are the main data collection device. These notes are detailed descriptions of settings, activities, and people. The observer records or paraphrases conversations. He may note individual characteristics, idiosyncracies, and non-verbal behavior. In addition, he writes up his tentative interpretation of what he sees, including its significance: whether it suggests hypotheses, and whether it confirms or contradicts other data or earlier hunches. As we saw in the example above, field notes are extensively quoted in the evaluation report to substantiate conclusions and recommendations.

Data analysis in ethnographic evaluations resembles the analytic procedures of investigative evaluation. As data collection takes place, the observer or other staff members read through field notes trying to spot trends, categorize the data, and generate hypotheses. The results of this analysis influence subsequent data collection, which in turn forms the basis for further analysis. As this process continues, the research becomes more focused, and ultimately the report framework emerges. The details of this process can vary substantially from investigator to investigator, which brings a very personal character to such qualitative research.

FURTHER CONSIDERATIONS

Perhaps the single most important decision in qualitative inquiry is the selection of the evaluator. How perceptive and credible the study is hinges on the skill, integrity, and perspective that the evaluator brings to it.

In a real sense, "the evaluator as instrument" represents both the great strength of these techniques and potentially their greatest problem. In reading a qualitative evaluation report, we should continually ask ourselves, "Whose view of the program are we really seeing?"

The controlling of personal bias is a vital concern in qualitative inquiry. Both the evaluator and the subjects of the study, teachers and children, are potential sources of bias. The evaluator's preconceived ideas of how things work and what are good and poor eduational practices can so distort what he sees and hears that what he reports may be little related to what happened. Similarly, subjects can produce biases and errors, either because they do not want to cooperate or because they do not know important information. They can bias observations by "masking" their behavior from the evaluator--that is, behaving differently because an observer is present. The professional qualitative evaluator is aware of this and attempts to safeguard against it in conducting the research. Persons of considerable skill and experience are required here, and qualitative evaluation should not be attempted without them.

NOTES ON SOURCES OF FURTHER INFORMATION

There are several general sources on qualitative research, including: Denzin (1970), Filstead (1970), Guba (1978), Lofland (1971), and Patton (1980).

Documentation. Carini (1975) and Perrone (1977) discuss the rationale for documentation. Perrone, in particular, looks at strengths and weaknesses of this approach. Engel (1975, 1977), Hein (1975), Burry (1978), and Suarez and Vandivere (1978) discuss how to go about documenting programs. Each author's work is aimed at different programs and audiences. For example, Hein looks at open education, and Suarez and Vandivere treat programs for preschool handicapped children.

Investigative evaluation. Because this is a new approach to evaluation, there are fewer sources and examples than for other methods. Murphy (1980) is a good guide for carrying out this kind of evaluation. Douglas (1976) discusses investigation methods, but because his research interests are far removed from classrooms, his book may be less useful than Murphy's. The Report Manual from the California Office of the Auditor General (1978) presents another perspective on investigative methods. Bissell (1979) also discusses this procedure, and reports by her and her staff are good examples of one form of investigative evaluation (e.g., Bissell, Potter, Barber, & Sheperd, 1978; Bissell, Potter, Herdell, & Tamayo, 1979).

Ethnographic evaluation. There is a large body of writings on educational anthropology and ethnography. Less has been written on applying these approaches to evaluation. General works of some applicability are Bruyn (1966), Good and Brophy (1978), Schatzman and Strauss (1973), McCall and Simmons (1969), and Rist (1975). Authors who discuss evaluation applications are Smith (1979), Stake (1978a), Berliner and Tikunoff (1977), and Jackson (1974). Wolcott (1975) and Mulhauser (1975) point to some problems in applying these methods to evaluation. Several good examples of this research exist that are evaluations or are much like evaluations; see Jackson (1968), Stake (1978b), Tikunoff, Berliner, and Rist (1975), Smith and Geoffrey (1968), and Rist (1973).

V. DEVELOPING AN EVALUATION PLAN

COMPARING THE METHODS

We have introduced in this resource book several evaluation techniques that hold considerable promise for generating information that can be useful to LEA staff in attempting to improve their programs. Table 4 provides a quick overview of the methods discussed and some of their distinctive characteristics. They vary considerably in the kinds of question they address, in the likely audience for the results, in the data collection methods emphasized, and in the human and physical resource requirements.

Although all the methods can address a wide range of questions, each is better suited to some questions than to others. Self-study and outside review typically focus on how well the program measures up to some set of standards or expectations of program performance. Structured process evaluation seems best suited to answering straightforward implementation questions, such as whether new curricular material is used in classrooms. Qualitative methods can investigate quite complex questions about human behavior and interaction in school settings.

As a result of the difference in emphasis, the audience for each method will also vary. The self-study report is aimed primarily at the program staff. Information from structured process evaluation may be most useful to those program administrators who have little direct contact with their program and desire some basic information about "what's going on." Outside review and investigative evaluations produce information that school boards, PACs, and program policy groups are likely to find most interesting.

As for data collection, although many of the methods use some form of observation, interviews, and document examination, each tends to rely on some

Table 4 Comparison of ECT-I

Improvement Method	Brief Description of Method	Human Resource Requirements
Self-Study	Committees of the staff compare aspects of program to some set of standards or expectation of the program in order to identify strengths and weaknesses	Teachers and other staff members collect and analyze data and write reports
Outside Review	Visiting committee examines the program to help ECT-I staff identify program strengths and weaknesses	Staff from nearby programs and other early childhood specialists collect and analyze data
Structured Process Evaluation	Evaluator assesses relatively straightforward implementation concerns such as whether materials are in classrooms; quite useful for administrators of large programs who may have infrequent contact with classrooms	Staff members fill out questionnaires and respond to interviews; local evaluator or program assistant collects and analyzes data
Product Evaluation	Evaluator examines tests and other outcome data to identify program strengths and weaknesses	Staff members collect test data; program director or evaluator analyzes data
Process-Product Evaluation	Evaluator investigates statistical relationships between various program processes and outcomes	Teachers and administrators may supply some information; observers and interviewers collect process data; statistician or evaluator analyzes data and interprets results
Documentation	Documentor collects, arranges, and interprets program documents to depict the program for the staff, administrators, and parents	Teachers, aides, parents, and administrators may collect program documents; trained documentor directs the evaluation
Investigative Evaluation	Evaluator uses interviews and qualitative techniques to investigate implementation and compliance questions for bodies such as school boards, state legislature, and Parent Advisory Committees	Highly trained outside evaluator conducts the study
Ethnographic Evaluation	Ethnographer examines fine-grained interactions using anthropological and ethnomethodological perspectives	Highly trained qualitative investigator conducts the evaluation

Program Improvement Strategies

Primary Data Collection Techniques	Time Considerations	Additional Considerations
Document examination; brief interviews; questionnaires	Time-consuming--may require several months	Can be, and often is, used to inform an outside review
Document examination; checklists; rating scales; informal interviews	Relatively quick by itself, but it can be time consuming if conducted in conjunction with other efforts such as self-study	A chief benefit may be dissemination of ideas that outside reviewers bring to and take home from the program
Structured observation; structured interviews	Relatively quick	Can be a very useful program management tool and is easy to implement
Norm- and criterion-referenced tests	Relatively quick assuming that a short-term impact evaluation is already being done	Helps in identifying where to make improvements but provides little information on how
Structured observation; structured interviews; questionnaires; standardized tests	Relatively time-consuming	Explicitly ties program attributes to outcomes; requires standardized measurement and therefore may inadvertently focus attention on easily measured but less important program variables
Document examination; self-reports; interviews; questionnaires; observation	Relatively quick	Depends on available records, whose quality may vary greatly
Intensive interviews; document examination; transient observation	Relatively quick	A new approach, as yet rarely applied to program evaluation; may raise ethical problems
Interviews; long-term observation; document examination	Relatively time-consuming	The description and analysis this method allows is a real strength; availability of a qualified

data collection procedures more than on others. For example, investigative evaluation uses primarily intensive interviewing, while ethnographic evaluation depends chiefly on observation; but the former also uses observations, and the latter, interviews. The choice of data collection procedures in part determines human resource requirements. Thus self-study, outside review, structured process evaluation, and product evaluation, which all use relatively structured procedures, can be done by persons who have little formal evaluation training. On the other hand, investigative evaluation, documentation, and ethnographic evaluation, which depend on open-ended techniques, require highly trained researchers.

Finally, the time needed for evaluation also differentiates the methods. Some approaches can provide information in a few weeks or months; others can take a year or more. Outside reviewers, for example, can examine a program in one week or less and return their suggested modifications a month later. On the other hand, an ethnographic study of a large program could take a year for data collection and another year for analysis.

GETTING STARTED: SELECTING AN APPROACH

As we have tried to illustrate throughout the resource book, no single evaluation method is always most appropriate. Rather, trade-offs between strengths and weaknesses make each more appropriate in some situations than in others. In selecting an approach, a first consideration is simply resources. If few resources are available for evaluation, or if most evaluation resources are allocated to other activities, methods such as process-product studies and ethnographic evaluation which tend to be resource intensive may not be feasible.

Moving beyond the resources issues, the next question is, "What is it that you want to know about the program?" Are you interested in examining

the whole program or just selected features, or even a single topic? An answer here usually starts with some impressions about problems with current practice, and some hunches or hypotheses about their causes. Discussions with a professional program evaluator will usually result in some suggestions for appropriate evaluation methods. For example, a program administrator may be troubled because a new reading curriculum doesn't seem to be working. She may wonder whether program materials have reached the classrooms, and whether teachers are using them regularly. A structured process evaluation probably can answer these questions. Evaluators can investigate how many teachers have received program materials, and when; and can observe classrooms to see whether the materials are used or are still in boxes.

In addition to technical considerations of matching an appropriate evaluation method to specific research questions, there are very important questions of personal interaction. In particular, the perspectives and interests of the target audiences should play a key role in developing an evaluation plan. If an audience is to find the results of an evaluation useful, it must view the research questions and proposed procedures as appropriate, and have confidence in the credibility and integrity of the evaluators. The involvement of these "stakeholders" throughout the entire evaluation, from developing the initial plan to conducting the study and reviewing the preliminary findings, seems crucial in engaging their commitment to the evaluation and their use of the findings.

IMPLEMENTING THE PLAN

Evaluation efforts, like programs themselves, can be quite complex, and as a result are sometimes difficult to carry out successfully. Some methods entail the collaboration of several people over an extended time. For example,

self-study may require the part-time participation of the whole staff for several months. Other methods demand intricate data collection procedures. In particular, investigative evaluation, documentation, and ethnographic evaluation depend upon careful interviewing, document review, and observation. Still others, such as process-product studies, involve complex statistical data analysis.

Staff motivation is, of course, critical. When staff members are fully involved as stakeholders in the evaluation they are likely to take the effort seriously. If, on the other hand, evaluation is mandated from outside the program, or if a directive comes down that the staff will participate, evaluation by any method is likely to be more difficult. Clearly, motivation becomes more important in methods that require staff members to collect or analyze data. Thus, self-study may fail if the staff believes it is a waste of time. More generally, if staff members fear the evaluation, see it as intrusive, or think it is unneeded, it seems unlikely that any method will be successful.

The availability of appropriate evaluation expertise is also critical. As we have pointed out, several methods require ingenuity, intelligence, training, and experience on the part of the people performing them. Clearly the quality of ethnographic and investigative evaluation, for example, will vary enormously depending upon who conducts the study. Similarly, just designating someone as the "project documentor" will hardly ensure that a good, or even a sufficient, documentation will result.

CONCLUDING REMARK

While keeping this resource book brief, we also have tried to make it rich in ideas about alternative evaluation methods and to illustrate how

their use might ultimately lead to program improvement. What we have presented here, however, is only a sample of what is possible. Further, the methods outlined above are not in any sense rigidly defined, but rather general approaches that must be tailored to a variety of technical and human considerations. An underlying theme throughout is that there is a great potential here which few LEAs have yet tapped. We hope that this resource book provides some encouragement in moving in that direction.

REFERENCES

Airasian, P. W., and Madaus, G. F. Criterion-referenced testing in the classroom. Measurement in Education, 1972, 3, 1-7.

APA, AERA, & NCME. Standards for educational and psychological tests (Reprinted in Buros, 1974). Washington, D.C.: APA, 1974.

Arkansas Elementary School Council. Self evaluation. A guide for elementary schools of Arkansas. Little Rock, Ark.: Arkansas State Department of Education, 1971.

Baker, E. L., Formative evaluation of instruction. In W. J. Popham (Ed.). Evaluation in education. Berkeley, Ca.: McCutchan Publishing Co., 1974.

Behnke, G., Bennett, J., Chase, C., Day, J., Lazar, C., and Mittleholtz, D. Coping with classroom distractions. San Diego: San Diego Interactive Research and Development Team, no date.

Berliner, D. C., and Rosenshine, B. The acquisition of knowledge in the classroom. Technical Report IV-I. San Francisco, Ca.: Far West Regional Laboratory, 1976.

Berliner, D. C., and Tikunoff, W. Ethnography in the classroom. In G. D. Borich (Ed.). The appraisal of teaching: Concepts and process. Reading, Ma.: Addison-Wesley, 1977.

Bissell, J. S., Use of evaluation and policy studies by the California legislature. Educational Evaluation and Policy Studies, 1979, 1, 29-37.

Bissell, J. S., Christophel, R. E., Sequira, D. L., and Farias, E. G. Bilingual education: Pupil assessment, program evaluation and local program implementation. Report P-828.2. Sacramento, Ca.: Office of the Auditor General, 1979.

Bissell, J. S., Potter, E. T., Barber, J. M., and Sheperd, E. H. California's policies for education of gifted pupils need legislative redefinition. Report 820. Sacramento, Ca.: Office of the Auditor General, 1978.

Bissell, J. S., Potter, E. T., Herdell, K., and Tamayo, A. Improvements needed in award management of childcare contracts. Report P-857.1. Sacramento, Ca.: Office of the Auditor General, 1979.

Boehm, A. E., and Weinberg, R. A. The classroom observer. A guide for developing observation skills. New York: Teachers College Press, 1977.

Bruyn, S. The human perspective in sociology: The methodology of participant observation. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Burges, B. Facts for a change: Citizen action research for better schools. Boston, Ma.: Institute for Responsive Education, 1976.

Burry, J. Documentation strand: teacher corps national development conference. Los Angeles: UCLA-CSE, 1978.

Cannell, C., and Kahn, R. The collection of data by interviewing. In L. Festinger and D. Katz (Eds.). Research methods in the behavioral sciences. New York: Holt, Rinehart and Winston, 1953.

Carini, P. F. Observation and description: An alternative methodology for the investigation of human phenomena. Grand Forks, N.D.: North Dakota Study Group on Evaluation, 1975.

Cassell, J. A fieldwork manual for studying desegregated schools. Washington, D.C.: The National Institute of Education, 1978.

Cooley, W. W. Exploratory observational studies. Educational Researcher, 1978, 7, 9-16.

Denzin, N. The research act. Chicago, Ill.: Aldine Publishing Co., 1970.

Douglas, J. D. Investigative social research: Individual and team field research. Beverly Hills, Ca.: Sage Publishing Co., 1976.

Engel, B. S. A handbook on documentation. Grand Forks, N.D.: North Dakota Study Group on Evaluation, 1975.

Engel, B. S. Informal evaluation. Grand Forks, N.D.: North Dakota Study Group on Evaluation, 1977.

Fenstermacher, G. D. Remarks to the board of education. San Diego Unified School District, February 10, 1977.

Ferreira, J., and Burges, B. Collecting evidence: A layman's guide to participant observation. Boston, Ma.: Institute for Responsive Education, 1976.

Filstead, W. J. (Ed.). Qualitative methodology. Chicago, Ill.: Markham, 1970.

Flanders, N. A. Teacher influence, pupil attitudes and achievement. Washington, D.C.: U.S. Office of Education (Cooperative Research Monograph No. 12), 1965.

Georgiades, W. How good is your school? Reston, Va.: The National Association of Secondary School Principals, 1978.

Good, T.L., and Brophy, J.E. Looking in classrooms, 2nd ed. New York: Harper and Row, 1978.

Guba, E. G. Toward a methodology of naturalistic inquiry in educational evaluation. Los Angeles, Ca.: Center for the Study of Evaluation, 1978.

Hall, G. E., and Loucks, S. F. A developmental model for determining whether the treatment is actually implemented. American Educational Research Journal, 1977, 14, 236-276.

Haney, W. Short-term impact evaluations of early childhood Title I programs. Cambridge, Ma.: The Huron Institute, 1980.

Haney, W., and Gelberg, W. Assessment in early childhood education. Cambridge, Ma: The Huron Institute, 1980.

Hein, G. E. An open education perspective on evaluation. Grand Forks, N.D.: University of North Dakota Press, 1975.

Holley, F. M. What it takes to win: Factors in the utilization of evaluation findings for educational improvement. Austin, Tx.: Austin Independent School District, 1978.

Isaac, S. A statewide study of ECE and non-ECE schools on reading achievement, grades 2 and 3, based on the California assessment program results. Studies in Educational Evaluation, 1977, 3, 179-194.

Jackson, P. W. Life in classrooms. New York: Holt, Rinehart and Winston, 1968.

Jackson, P. W. Naturalistic studies of school and classrooms. In M. W. Apple, M. J. Subkovick, and H. W. Lufler, Jr. (Eds.). Educational evaluation: Analysis and responsibility. Berkeley, Ca.: McCutchan Publishing Co., 1974.

Kennedy, M. Longitudinal information systems in early childhood Title I programs. Cambridge, Ma.: The Huron Institute, 1980.

Kerlinger, F. N. Foundations of behavioral research, 2nd ed. New York: Holt, Rinehart and Winston, 1973.

Lofland, J. Analyzing social settings. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Maccoby, E. and Maccoby, N. The interview: A tool of social science. In G. Lindzey (Ed.). Handbook of social psychology. Reading, Ma.: Addison-Wesley, 1954.

McCall, G. J., and Simmons, J. L. Issues in participant observation. Reading, Ma.: Addison-Wesley, 1969.

Medley, D., and Mitzel, H. Measuring classroom behavior by systematic observation. In N. Gage (Ed.). Handbook of research on teaching. Skokie, Ill.: Rand McNally, 1963.

Millman, J. Criterion-referenced measurement. In W. J. Popham (Ed.). Evaluation in education. Berkeley, Ca.: McCutchan Publishing Co., 1974.

Morris, L. L., and Fitz-Gibbon, C. T. How to measure program implementation. Beverly Hills, Ca.: Sage Publishing Co., 1978.

Mulhauser, F. Ethnography and policy making: The case of education. Human Organization, 1975, 34, 311-315.

Murphy, J. T. Getting the facts: A fieldwork guide for evaluation and policy analysis. Santa Monica, Ca.: Goodyear Publishing Co., 1980.

National Council for Accreditation of Teacher Evaluation. Standards for accreditation of teacher evaluation. Washington, D.C.: Author, 1970.

National Study of School Evaluation. Elementary school evaluation criteria. Arlington, Va.: Author, 1973.

Office of the Auditor General. Report manual. Sacramento, Ca.: Author, 1978.

Patton, M.Q. Qualitative evaluation methods. Beverly Hills, Ca.: Sage Publishing Co., 1980.

Payne, S. L. The art of asking questions. Princeton, N.J.: Princeton University Press, 1951.

Pedhazur, E. J. Analytic methods in studies of educational effects. In F. N. Kerlinger (Ed.). Review of research in education, vol. III. Itasca, Ill.: F. E. Peacock Publishers, Inc., 1975.

Perrone, V. Documentation: A process for classroom/program evaluation and personal/professional learning. University of North Dakota, 1977 (ERIC No. ED 129684).

Rist, R. C. Ethnographic techniques and the study of an urban school. Urban Education, 1975, 1, 86-107.

Rist, R. C. The urban school: A factor for failure. Cambridge, Ma.: MIT Press, 1973.

Rosenshine, B., and Furst, N. The use of direct observation to study teaching. In R. Travers (Ed.). Second handbook on research on teaching. Skokie, Ill.: Rand McNally, 1973.

Rossi, P. H., Freeman, H. E., and Wright, S. R. Evaluation: A systematic approach. Beverly Hills, Ca.: Sage Publishing Co., 1979.

Roupp, R., Travers, J., Glantz, F., and Coelen, C. Children at the center: Summary findings and their implications, Part I (Volume 1 of Final Report of the National Day Care Study). Cambridge, Ma.: Abt Associates, 1979.

Sax, G. The use of standardized tests in evaluation. In W. J. Popham (Ed.). Evaluation in education. Berkeley, Ca.: Sage Publishing Co., 1974.

Schatzman, L., and Strauss, A. Field research: Strategies for a natural sociology. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Scriven, M. The evaluation of teachers and teaching. In G. D. Borich (Ed.). The appraisal of teaching: Concepts and process. Reading, Ma.: Addison-Wesley, 1977.

Scriven, M. The methodology of evaluation. Lafayette. In.: Social Science Education Consortium, 1966.

Sellitz, C., Wrightman, L.S., and Cook, S.W. Research methods in social relations. New York: Holt, Rinehart, & Winston, 1976.

Smith, L. M. An evolving logic of participant observation, educational ethnography and other case studies. In L. Schulman (Ed.). Review of research in education. Chicago: Peacock Press, 1979.

Smith, L. M., and Pohland, P. A. Education technology and the rural highlands. In Sjogren, D. (Ed.). Four evaluation examples: Anthropological, economic, narrative, and portrayal. Chicago: Rand McNally, 1974, pp. 5-54.

Smith, L. M., and Geoffrey, W. The complexities of an urban classroom. New York: Holt, Rinehart and Winston, 1968.

Stake, R. E. The case study method in social inquiry. Educational Researcher, 1978, 5-8. (a)

Stake, R. E. PLATO and fourth grade mathematics. Journal of Research on Children's Mathematics, 1978, 2 (2). (b)

Stallings, J. A. Learning to look: A handbook on classroom observation and teaching methods. Belmont, Ca.: Wadsworth Publishing Co., 1977.

Suarez, T. M., and Vandivere, P. Planning for evaluation. A resource book for programs for preschool handicapped children: Documentation. Chapel Hill, N.C.: The Technical Assistance Development System, 1978.

Tallmadge, K., and Wood, C. User's guide: ESEA Title I evaluation and reporting system (revised). Mountain View, Ca.: RMC Research Corporation.

Tikunoff, W., Berliner, D. C., and Rist, R. C. An ethnographic study of the forty classrooms of the beginning teachers evaluation study: A known sample (Technical Report #75-10-10-5). San Francisco, Ca.: Far West Regional Laboratory, 1975.

Travers, R. M. W. An introduction to educational research (Fourth edition). New York: Macmillan Publishing Co., 1978.

Tuckman, B. W. Conducting educational research. New York: Harcourt, Brace, Jovanovich, Inc., 1972.

Walberg, H. J. Educational process evaluation. In M. W. Apple, M. J. Subkoviek, and H. S. Lufler, Jr. (Eds.). Educational evaluation: Analysis and responsibility. Berkeley, Ca.: McCutchan Publishing Co., 1974.

Weick, K. Systematic observational methods. In G. Lindsey and E. Aronson (Eds.). The handbook of social psychology. (Second edition). Reading, Ma.: Addison-Wesley, 1968.

Weiss, C. H. Evaluation research. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Wolcott, H. Criteria for an ethnographic approach to research in schools. Human Organization, 1975, 34, 111-125.

Wolf, K. M. Evaluation in education. New York, N.Y.: Praeger, 1979.